

## 3 Gramáticas libres de contexto

Siguiendo la línea trazada en el capítulo anterior, en el presente exploraremos el siguiente nivel en la jerarquía de Chomsky, las gramáticas de tipo II o gramáticas libres de contexto.

Los lenguajes regulares, a pesar de su gran utilidad, su restricción a patrones simples los vuelve muy limitados y poco adecuados para la descripción del lenguaje natural. Siguiendo el ejemplo en [3], un lenguaje de la forma  $\{a^i bc^i \mid i \in \mathbb{N}\}$  no es regular<sup>1</sup> y corresponde a la anidación de elementos, así oraciones del tipo "Si el pasto es verde, entonces el pasto es verde", "Si si el pasto es verde, entonces el pasto es verde, entonces el pasto es verde" que se describen con la gramática  $\{(Si)^i \text{ el pasto es verde, (entonces el pasto es verde)}^i \mid i \in \mathbb{N}\}$  no serían gramaticalmente válidas en el español, pero lo son (a pesar de su poca naturalidad).

Por ello, surge el estudio de los lenguajes libres de contexto, clásicamente definidos mediante gramáticas de nombre homónimo, árboles de derivación o por medio de autómatas con pila —autómatas que tienen memoria—.

Por la vía categórica, en este capítulo mostraremos que gramáticas definidas mediante multicategorías, categorías donde los morfismos tienen como dominio más de un objeto, tienen exactamente el mismo poder expresivo que los lenguajes libres de contexto.

### 3.1. Multicategorías

Como en el capítulo anterior, comenzaremos dando los ingredientes para describir una multicategoría libre.

**Definición 4.** Una **signatura de multicategoría**  $H$  es un conjunto de nodos  $H_1$  y uno de objetos  $H_0$  junto a un par de funciones

$$H_0^* \xleftarrow{\text{dom}} H_1 \xrightarrow{\text{cod}} H_0$$

Un morfismo de firmas de multicategorías  $\varphi : H \rightarrow \Gamma$  es un par de funciones

---

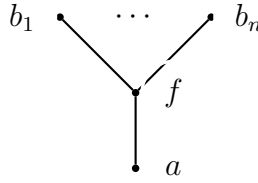
<sup>1</sup>Esta es una consecuencia inmediata del lema del bombeo para lenguajes regulares cuyo enunciado y prueba puede consultarse en [2]

$\varphi_0 : H_0 \rightarrow \Gamma_0$  y  $\varphi_1 : H_1 \rightarrow \Gamma_1$  tales que el siguiente diagrama conmuta:

$$\begin{array}{ccccc} H_0^* & \xleftarrow{\text{cod}} & H_1 & \xrightarrow{\text{dom}} & H_1 \\ \downarrow \varphi_0 & & \downarrow \varphi_1 & & \downarrow \varphi_0 \\ \Gamma_0^* & \xleftarrow{\text{cod}} & \Gamma_1 & \xrightarrow{\text{dom}} & \Gamma_1 \end{array}$$

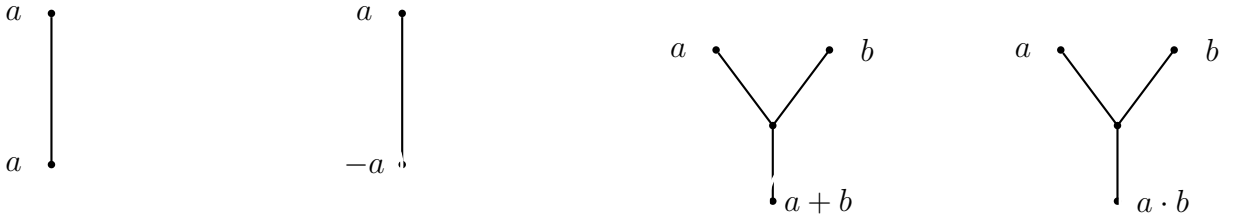
Con tales morfismos, las firmas de multicategorías forman la categoría **MultiSig**.

Dado una firma de multicategoría, denotaremos un nodo  $b_1 \dots b_n \xrightarrow{f} a$  gráficamente como:



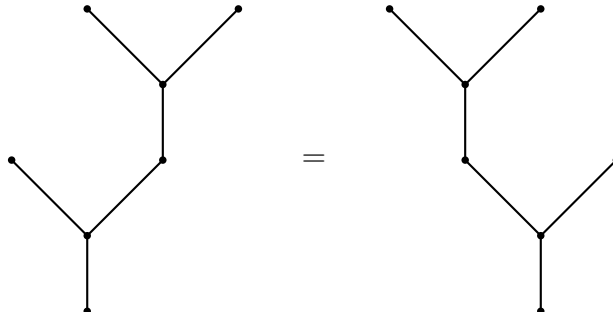
Además, diremos que  $f$  tiene aridad  $n$ .

**Ejemplo 5.** Consideremos la siguiente firma de multicategoría donde los objetos corresponderán a los números reales,  $\mathbb{R}$ . Luego, dado  $a, ab \in \mathbb{R}^*$  tenemos los siguientes nodos:



Notemos que los primeros dos nodos corresponden a las operaciones de aridad 1 que lleva un número asimismo o a su inverso aditivo respectivamente. Por otro lado, los últimos dos nodos corresponden a las operaciones binarias usuales de los números reales: la adición y la multiplicación.

Notemos, además, que para cualquier  $abc \in \mathbb{R}^*$  tenemos que, dado que nuestras operaciones son asociativas, entonces  $a + (b + c) = (a + b) + c$ , así gráficamente tenemos la igualdad de los siguientes diagramas:



Por ello, siguiendo un razonamiento inductivo para cualesquiera  $a_1 \dots a_n \in \mathbb{R}$  podemos agregar los nodos:



Así, todos los nodos son los representados por alguno de los diagramas anteriores. Con lo anterior, tenemos una signatura de multicategoría.

Ahora, veamos la definición principal de esta sección:

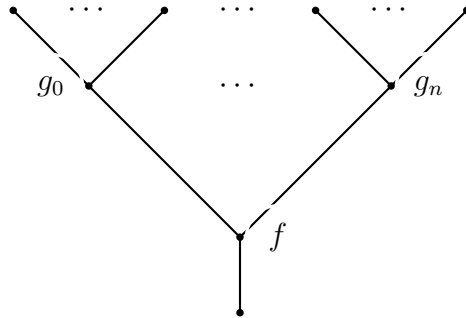
**Definición 5.** Un **multicategoría** es una signatura de multicategoría  $H$  junto a una operación de composición

$$\prod_{b_i \in \vec{b}} H_1(\vec{c}_i, b_i) \times H_1(\vec{b}, a) \rightarrow H_1(\vec{c}, a)$$

con  $a \in H_0$  y  $\vec{c}_i, \vec{b} \in H_0^*$  donde dadas  $f : \vec{b} \rightarrow a$  y  $g_i : \vec{c}_i \rightarrow b_i$  definimos la composición como:

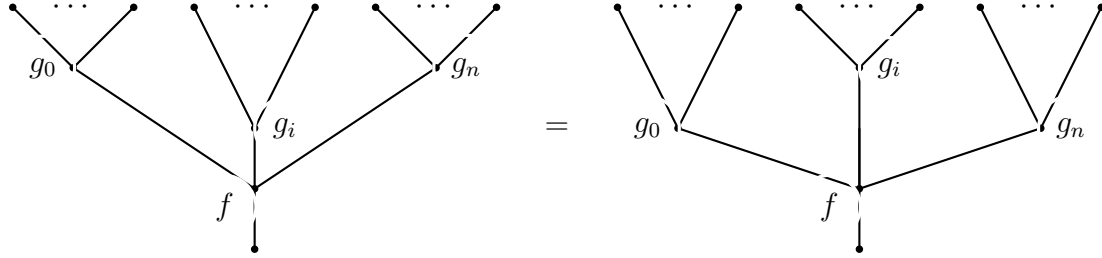
$$(c_1^1, c_2^1, \dots, c_{m_1}^1, c_2^1, \dots, c_1^n, \dots, c_{m_n}^n) \mapsto f(g_1(c_1^1, \dots, c_{m_1}^1), \dots, g_n(c_1^n, \dots, c_{m_n}^n))$$

que representamos gráficamente de la siguiente manera



Además, para cualquier  $a \in H_0$  tenemos una identidad  $a \xrightarrow{\text{Id}_a} a$  en donde se satisfacen las siguientes propiedades:

1. Para cualquier  $f : \vec{b} \rightarrow a$  en  $H_1$ ,  $f \cdot \text{Id}_a = f = \vec{\text{Id}}_{b_i} \cdot f$ ; y
2. Para cualesquiera  $f : \vec{b} \rightarrow a$  y  $g_i : \vec{c}_i \rightarrow b_i$  tenemos que:



Un **álgebra de firmas de multicategorías**  $F : \mathbf{H} \rightarrow \mathbf{N}$  es un morfismo de firma de multicategorías compatible con la composición, es decir, cada que  $\vec{c} \xrightarrow{\vec{g}} \vec{b} \xrightarrow{f} a$ ,  $F(\vec{g} \cdot f) = F(\vec{g}) \cdot F(f)$  Con estos morfismos, denotamos como **MultiCat** a la categorías de multicategorías.

Es importante resaltar que la propiedad 2 nos indica intuitivamente que no importa el orden en que hagamos las composiciones, es decir, se respeta cierta noción de asociatividad. Aunque todavía no definimos formalmente el significado del cálculo en diagramas, lo analizaremos con precisión en el siguiente capítulo.

La firma de multicategoría del ejemplo 5 es propiamente un multicategoría con la composición dada de la manera natural.

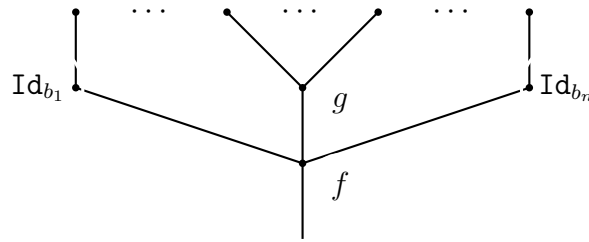
De manera análoga al capítulo anterior, nuestra intención a continuación es definir una multicategoría libre. Antes, veamos la siguiente definición

**Definición 6.** Sea  $H$  una firma de multicategoría. Sean  $f : \vec{b} \rightarrow a$  y  $g : \vec{c} \rightarrow b_i$  en  $H_1$  con  $i \in \{1, \dots, n\}$  con  $n$  es la aridad de  $f$ .

Definimos la **composición parcial en la  $i$ -ésima entrada** como:

$$g \circ_i f = (\text{Id}_{b_1} \cdots g \cdots \text{Id}_{b_n}) \cdot f$$

Gráficamente



Una observación importante es que definimos la composición parcial a partir de la composición, pero también pudimos definirlo inversamente de la siguiente forma:

$$[g_1 \cdots g_n] \cdot f = g_n \circ_n (g_{n-1} \circ_{n-1} \cdots \circ_2 (g_1 \circ_1 f) \cdots)$$

Donde la propiedad 2 de la composición para multicategorías nos garantiza que efectivamente la composición anterior coincide con la usual.

Con ello, ya estamos casi en condiciones de mostrar la construcción de una multicategoría libre, pero antes veamos una definición más.

**Definición 7.** Sea  $H$  una signature de multicategoría y sea  $a \in H_0$ , definimos recursivamente un **árbol con raíz  $a$  de  $H$** , como:

- Un nodo de la forma  $w : \varepsilon \rightarrow a$  es un árbol (a los nodos de esta forma, les llamamos hojas);
- Un par  $(f, (g_1, \dots, g_n))$  donde  $f$  es un nodo de aridad  $n$ ,  $f : a_1 a_2 \cdots a_n \rightarrow a$ , y  $g_1, \dots, g_n$  son nodos con codominios  $a_1, \dots, a_n$  es un árbol;
- Son todos.

Sea  $H$  una signature de multicategoría. **Multi**( $H$ ) es la multicategoría cuyos objetos son exactamente los mismos que  $H$  con los morfismos los árboles en  $H$  y para cada objeto  $a \in G_0$  tenemos la  $\text{Id}_a$ . La  $i$ -ésima composición parcial para morfismos  $f, g$  en **Multi**( $H$ ),  $g \circ_i f$  está por reemplazar la  $i$ -ésima hoja de  $f$  por una copia de  $g$ .

Decimos que dos árboles en **Multi**( $H$ ) son iguales si puede ser obtenido uno a partir de otro mediante la aplicación consecutiva de las propiedades 1 y 2 para la composición en multicategorías.

De acuerdo presentado previamente, la composición parcial anterior induce correctamente la composición y, por lo tanto, **Multi**( $H$ ) efectivamente es un multicategoría. Más aún, la construcción anterior es libre en virtud de la siguiente proposición.

**Proposición 4.** Sean  $H$  una signature de multicategorías,  $\mathbf{G}$  un multicategoría y  $f : G_1 \rightarrow O_1$  una función que preserve la aridad de los nodos. Entonces existe un único morfismo de multicategorías  $\varphi : \mathbf{Multi}(H) \rightarrow \mathbf{G}$  tal que el siguiente diagrama conmuta:

$$\begin{array}{ccc} H & \xrightarrow{\eta} & \mathbf{G} \\ \downarrow i & \nearrow \varphi & \\ \mathbf{Multi}(H) & & \end{array}$$

## 3.2. Lenguajes libres de contexto

Sea  $G$  una signature finita de multicategoría. Consideremos sus objetos como símbolos, los nodos como reglas de producción y los morfismos en **Multi**( $G$ ) como derivaciones, entonces obtenemos la noción de una gramática, más precisamente:

**Definición 8.** Un **gramática de multicategorías** es una signature finita de multicategoría de la siguiente forma:

$$(B + V)^* \xleftarrow{\text{dom}} G \xrightarrow{\text{cod}} B$$

donde  $V$  es un vocabulario y  $B$  es un conjunto de símbolos no terminales con un símbolo distinguido  $s \in B$  y  $G$  un conjunto de reglas de producciones.

El lenguaje generado por  $G$  es:

$$\mathcal{L}(G) = \{u \in V^* \mid \exists g : u \rightarrow s \in \mathbf{Multi}(G)\}$$

Para comprender la definición anterior, consideremos los siguientes ejemplos.

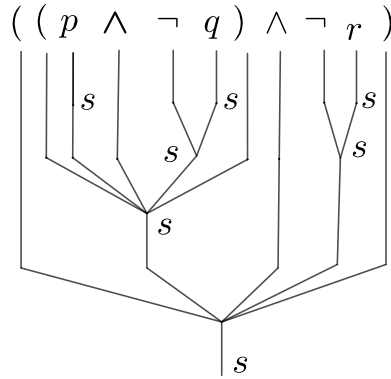
**Ejemplo 6.** El lenguaje de la lógica proposicional. Sea el conjunto de símbolos terminales vacío, el vocabulario  $V = Var \cup \{\wedge, \neg, (, )\}$  donde  $Var$  es el conjunto de letras proposicionales con las asignaciones:

$$\text{para todo } p \in Var, \quad p \rightarrow s$$

y las reglas de producción

$$(s \wedge s) \rightarrow s \quad \neg s \rightarrow s$$

Con lo anterior, podemos obtener, por ejemplo, la siguiente derivación



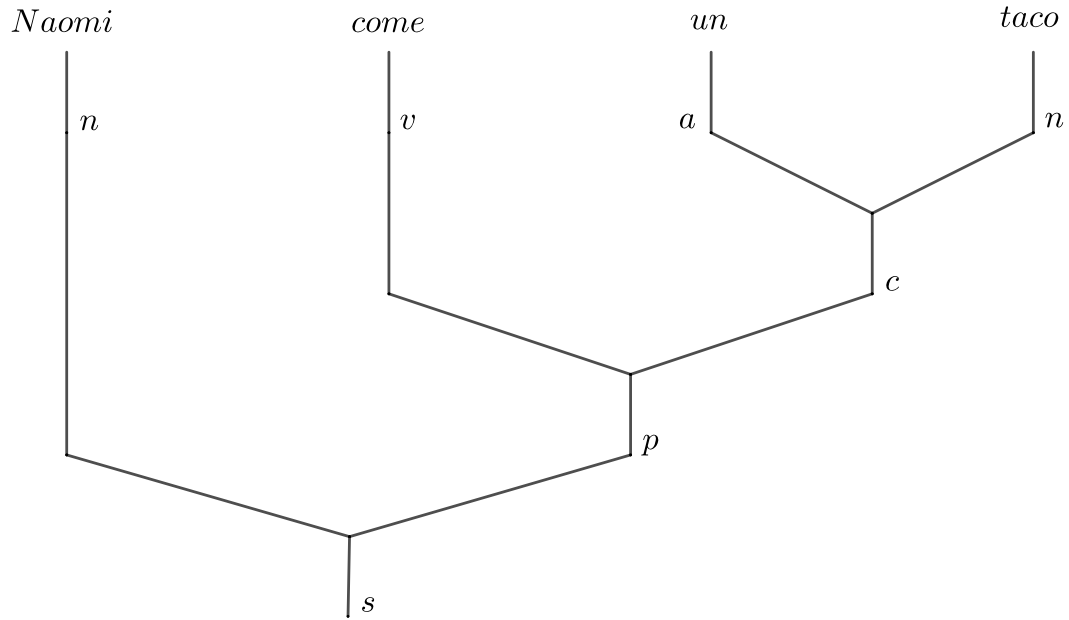
**Ejemplo 7.** Consideremos el siguiente conjunto de símbolos  $B = \{n, v, a, c, p\}$  que corresponde a sustantivo, verbo, artículo, complemento y predicado; el vocabulario  $V = \{Naomi, come, un, taco\}$ ; las asignaciones

$$Naomi \rightarrow n \quad come \rightarrow v \quad un \rightarrow n \quad taco \rightarrow n$$

y las reglas de producción

$$(a, n) \rightarrow c \quad (v, c) \rightarrow p \quad (n, p) \rightarrow s$$

Tenemos la siguiente derivación



Ahora, como anunciamos al inicio del capítulo, mostraremos la relación entre los lenguajes anteriores con la jerarquía de Chomsky. Para ello, veamos las siguientes definiciones y resultados.

**Definición 9.** Una gramática libre de contexto es una tupla  $G = (V, B, P, s)$  donde  $V$  es un vocabulario,  $B$  es un conjunto de símbolos no terminales con un símbolo distinguido  $s$  y  $P$  es un conjunto de reglas de producción de la forma  $A \rightarrow \alpha$  donde  $A$  es un símbolo no terminal y  $\alpha \in (V + B)^*$ .

Si  $A \rightarrow \alpha$  es una regla de producción, para cualquier  $\beta, \gamma \in (V + B)^*$  definimos la derivación  $\beta A \gamma \Rightarrow \beta \alpha \gamma$ . Consideramos  $\Rightarrow^*$  como la cerradura transitiva de  $\Rightarrow$ .

Definimos el lenguaje generado por  $G$  como:

$$\mathcal{L}(G) = \{w \in V^* \mid \exists s \Rightarrow w\}$$

Decimos que un lenguaje  $L$  es libre de contexto si existe una gramática  $G$  libre de contexto tal que  $L = \mathcal{L}(G)$

Observemos que una oración en un lenguaje libre de contexto corresponde a una cadena en  $V^*$  que puede ser derivada a partir de  $s$  en  $G$ .

Ahora, la manera anterior de describir el lenguaje es eficiente para la generación de nuevas expresiones, sin embargo, puede no ser muy útil para el propósito de discriminar si una expresión dada pertenece, o no, al lenguaje. Con ese propósito surgen los árboles de derivación.

**Definición 10.** Sea  $G = (V, B, P, s)$  una gramática libre de contexto.

Un árbol es un árbol de derivación para  $G$  si:

- Cada vértice está etiquetado con un símbolo de  $V \cup B \cup \{\varepsilon\}$ ;

- la raíz del árbol es  $s$ ;
- si un vértice interior está etiquetado con  $A$ , entonces  $A \in B$ , es decir, no es un símbolo terminal;
- si un vértice  $A$  tiene  $n$  hijos,  $A_1, \dots, A_n$ , entonces:

$$A \rightarrow A_1 A_2 \cdots A_n$$

es una regla de producción en  $P$ ; y

- si un vértice está etiquetado con  $\varepsilon$ , entonces es un hoja y, además, es el único hijo de su padre.

Notemos que el diagrama del ejemplo uno corresponde a un árbol de derivación en la gramática  $A \rightarrow \neg A | A \wedge A | A \vee A | A \rightarrow A | A \leftrightarrow A | a_0 | a_1 | a_2$ .

Con esto, podemos enunciar la siguiente proposición cuya prueba podemos consultar en [2].

**Proposición 5.** *Sea  $G = (V, B, P, s)$  una gramática libre de contexto. Entonces  $s \Rightarrow^* w$  si y solo si existe un árbol de derivación para  $G$  que produce a  $w$ .*

Dada la definición de morfismos en una multicategoría libre, es inmediato notar que los morfismos de la forma  $f : u \rightarrow s$  con  $u \in V^*$  en una gramática de multicategoría son árboles de derivaciones y viceversa. Así, tenemos el siguiente teorema con el cual terminamos el presente capítulo.

**Teorema 2.** *Sea  $G$  una signatura de multicategoría y  $G'$  una gramática libre de contexto, entonces*

1.  $L(G)$  es un lenguaje libre de contexto.
2. Existe una signatura de multigráfica que genera a  $L(G')$ .